FIG. 1

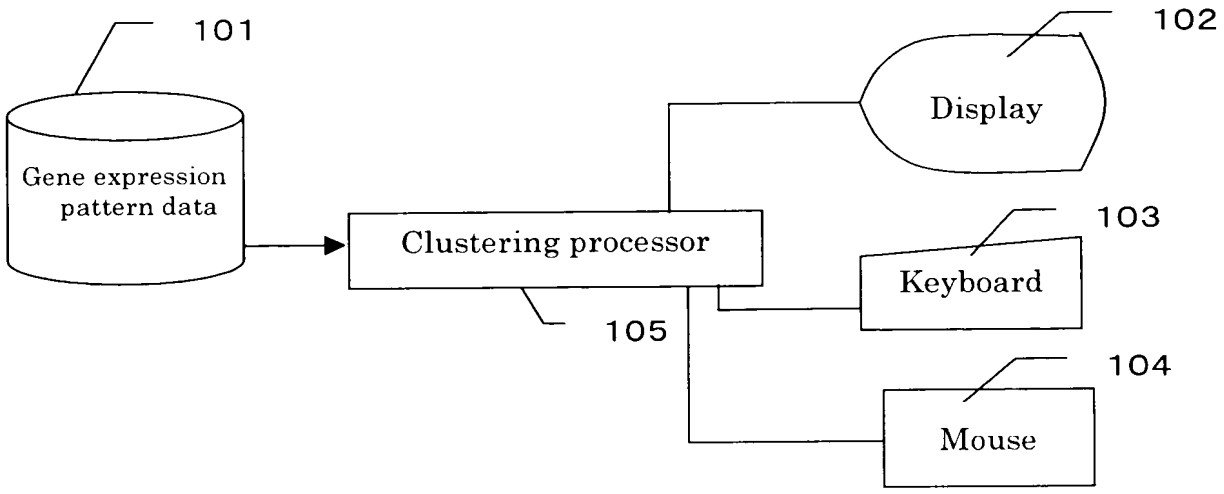

101
Gene expression pattern data
Clustering processor
105
102
Display
103
Keyboard
104
Mouse

FIG. 2

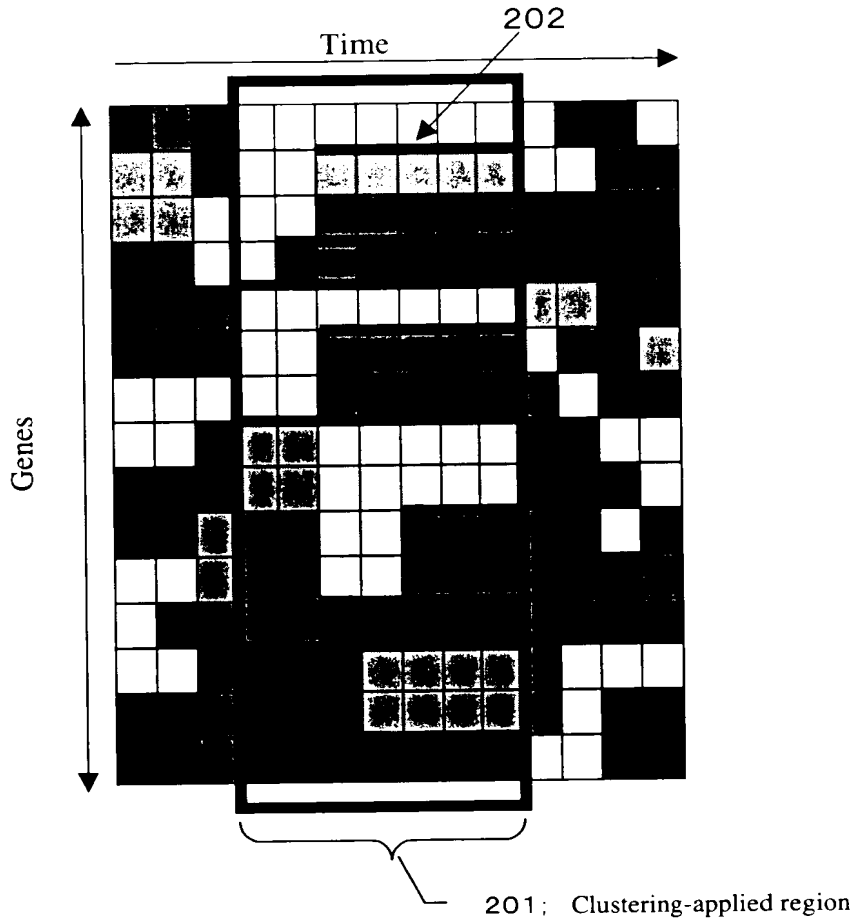


FIG.3

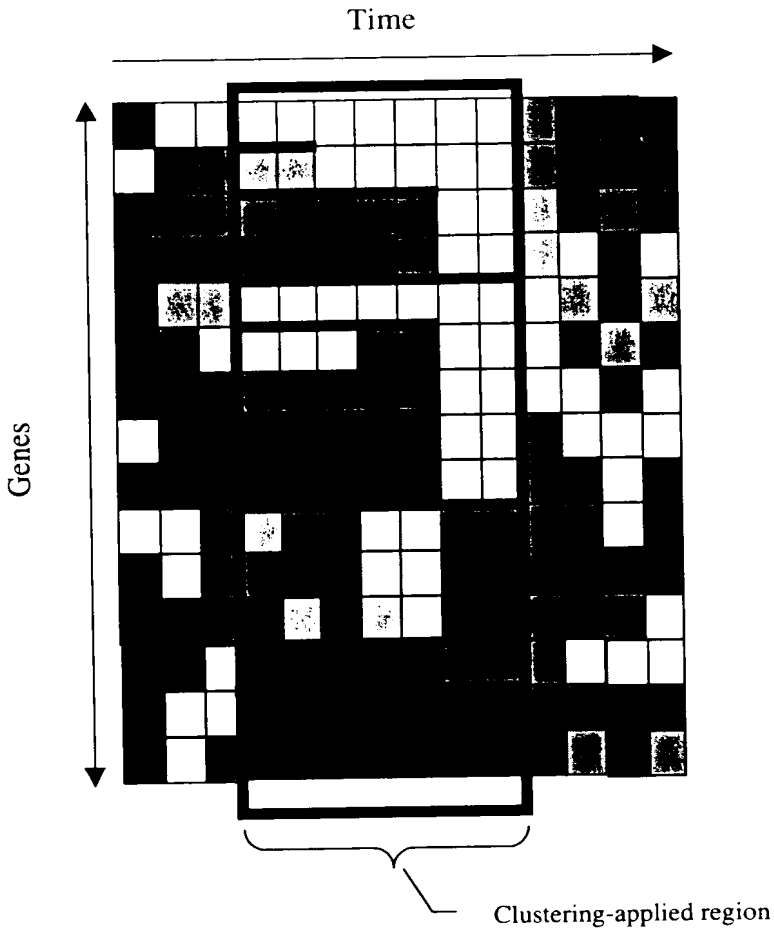


FIG. 4

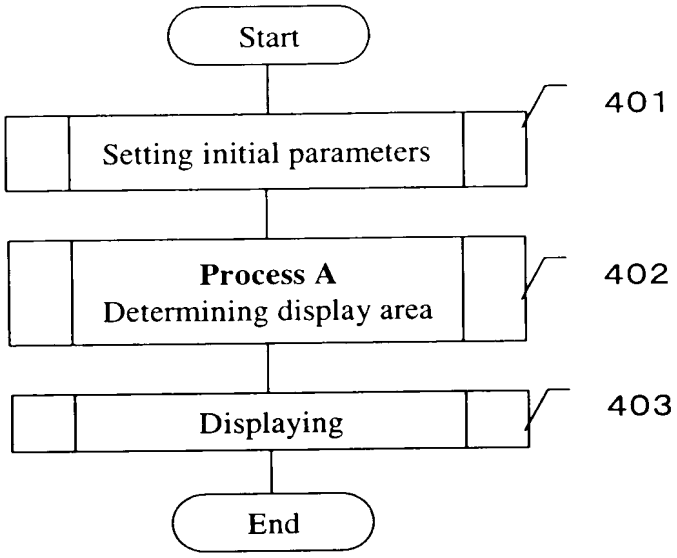

Start
Setting initial parameters
401
Process A
Determining display area
402
Displaying
403
End

FIG. 5

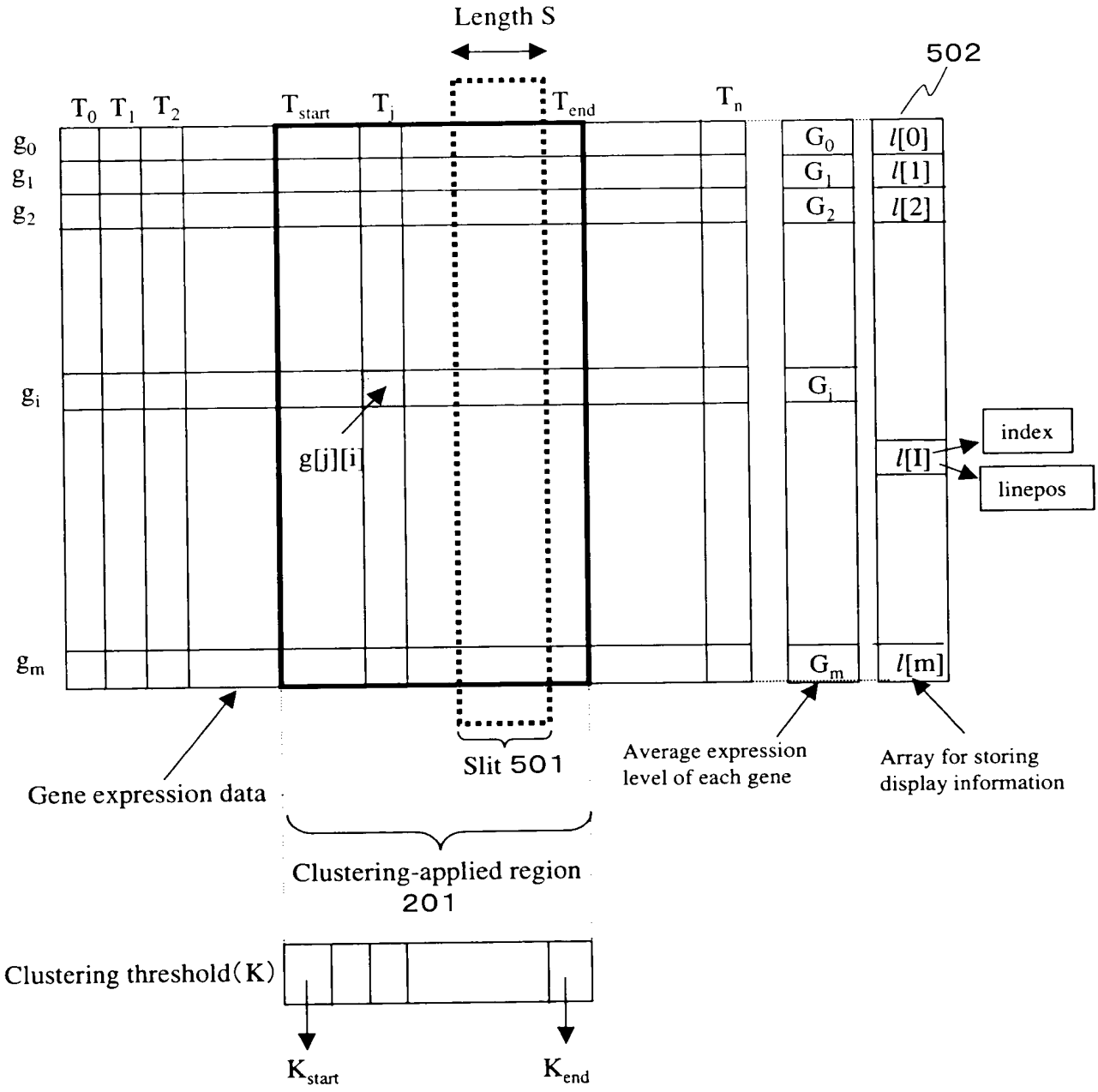

Length S
502
T_0 T_1 T_2 T_{start} T_j T_{end} T_n
g_0
g_1
g_2
g_i
g_m
g[j][i]
G_0
G_1
G_2
G_i
G_m
l[0]
l[1]
l[2]
l[I]
l[m]
index
linepos
Slit 501
Gene expression data
Average expression level of each gene
Array for storing display information
Clustering-applied region
201
Clustering threshold(K)
K_{start}
K_{end}

FIG. 6

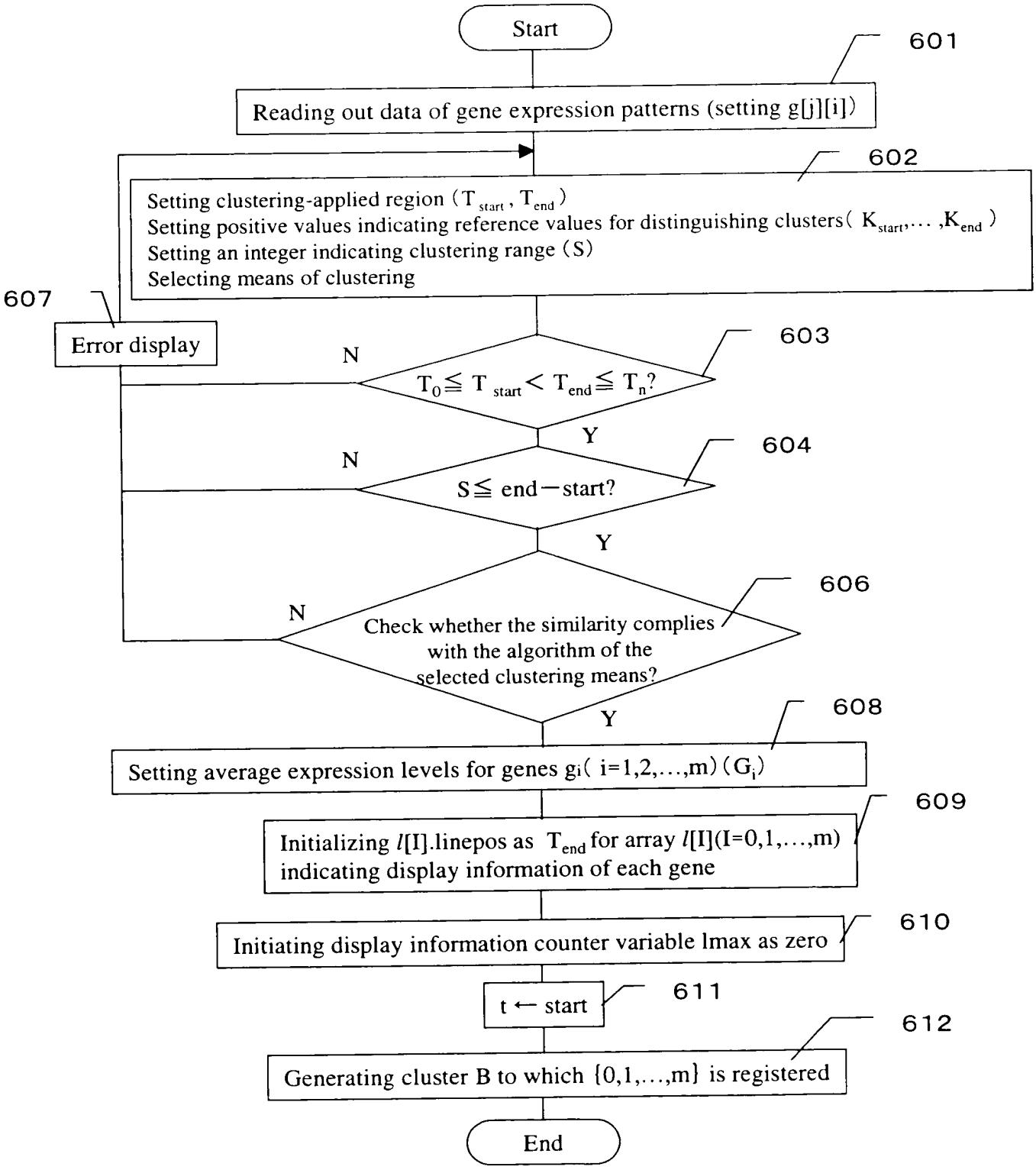

Start
601
Reading out data of gene expression patterns (setting g[j][i])
602
Setting clustering-applied region (T_{start}, T_{end})
Setting positive values indicating reference values for distinguishing clusters(K_{start},... ,K_{end})
Setting an integer indicating clustering range (S)
Selecting means of clustering
607
Error display
603
$T_0 \leqq T_{start} < T_{end} \leqq T_n$?
N
Y
604
$S \leqq$ end − start?
N
Y
606
Check whether the similarity complies with the algorithm of the selected clustering means?
N
Y
608
Setting average expression levels for genes g_i(i=1,2,...,m) (G_i)
609
Initializing l[I].linepos as T_{end} for array l[I](I=0,1,...,m) indicating display information of each gene
610
Initiating display information counter variable lmax as zero
611
t ← start
612
Generating cluster B to which {0,1,...,m} is registered
End

FIG. 7

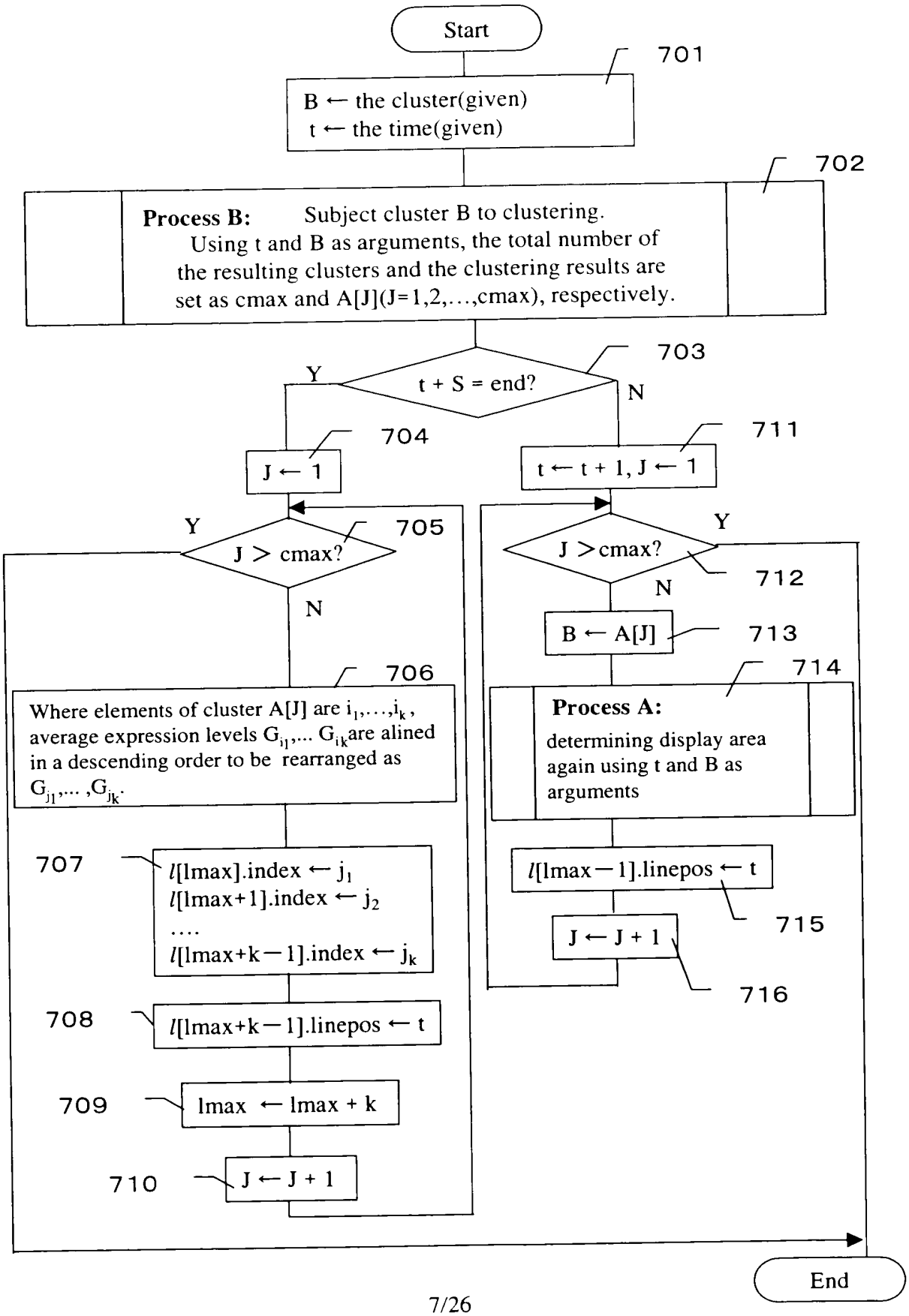

Start
701
B ← the cluster(given)
t ← the time(given)
702
Process B: Subject cluster B to clustering.
Using t and B as arguments, the total number of the resulting clusters and the clustering results are set as cmax and A[J](J=1,2,...,cmax), respectively.
703
t + S = end?
Y
N
704
J ← 1
711
t ← t + 1, J ← 1
705
J > cmax?
Y
N
712
J > cmax?
Y
N
706
Where elements of cluster A[J] are $i_1,\ldots,i_k$, average expression levels $G_{i1},\ldots G_{ik}$ are alined in a descending order to be rearranged as $G_{j1},\ldots,G_{jk}$.
707
l[lmax].index ← j_1
l[lmax+1].index ← j_2
....
l[lmax+k−1].index ← j_k
708
l[lmax+k−1].linepos ← t
709
lmax ← lmax + k
710
J ← J + 1
713
B ← A[J]
714
Process A:
determining display area again using t and B as arguments
l[lmax−1].linepos ← t
715
J ← J + 1
716
End

FIG. 8

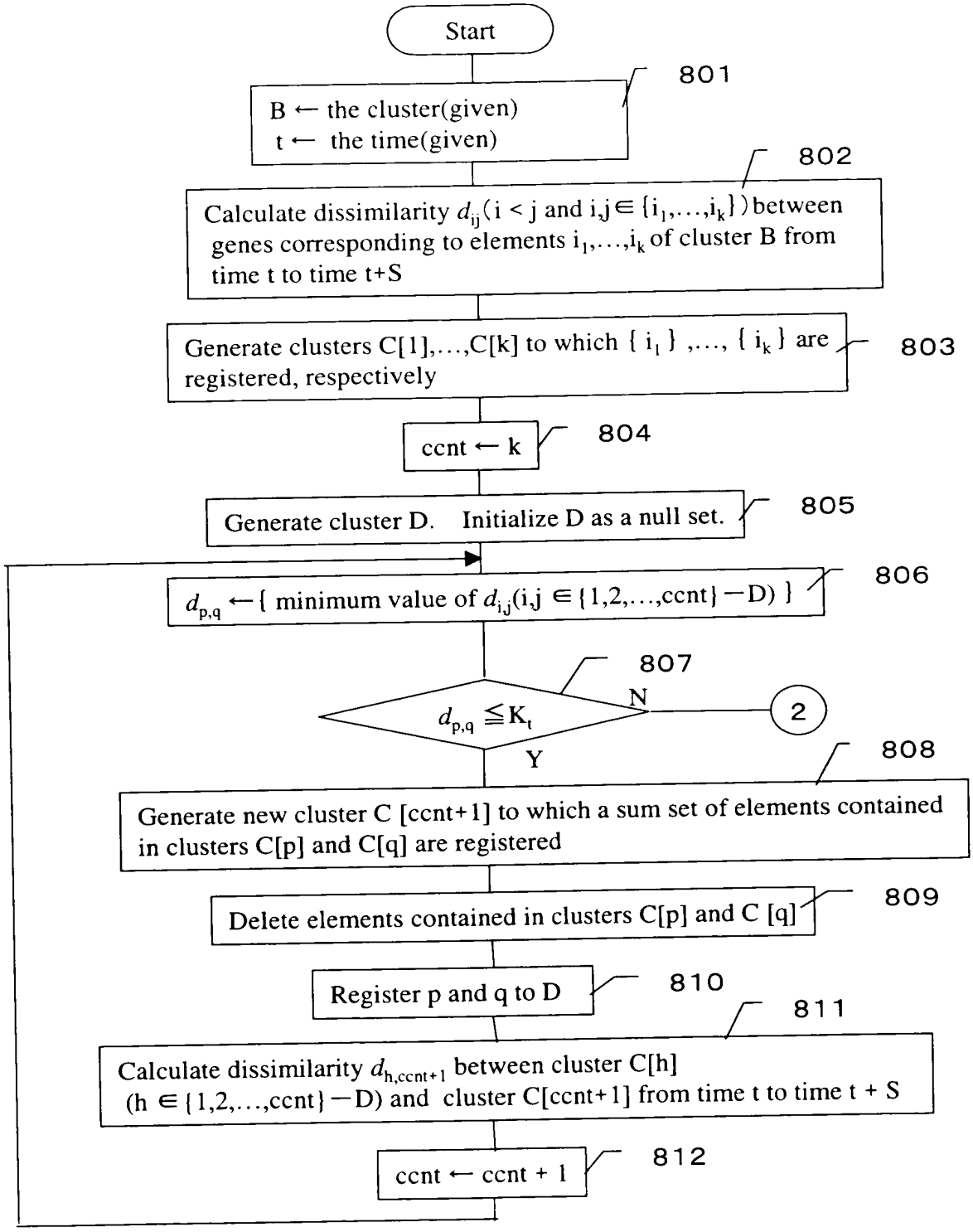

Start
801
B ← the cluster(given)
t ← the time(given)
802
Calculate dissimilarity d_{ij} (i < j and i,j ∈ {i_1,…,i_k}) between genes corresponding to elements i_1,…,i_k of cluster B from time t to time t+S
Generate clusters C[1],…,C[k] to which { i_1 } ,…, { i_k } are registered, respectively
803
ccnt ← k
804
Generate cluster D. Initialize D as a null set.
805
$d_{p,q}$ ← { minimum value of $d_{i,j}$(i,j ∈ {1,2,…,ccnt} − D) }
806
807
$d_{p,q} \leqq K_t$
N
2
Y
808
Generate new cluster C [ccnt+1] to which a sum set of elements contained in clusters C[p] and C[q] are registered
Delete elements contained in clusters C[p] and C [q]
809
Register p and q to D
810
811
Calculate dissimilarity $d_{h,ccnt+1}$ between cluster C[h] (h ∈ {1,2,…,ccnt} − D) and cluster C[ccnt+1] from time t to time t + S
ccnt ← ccnt + 1
812

FIG. 9

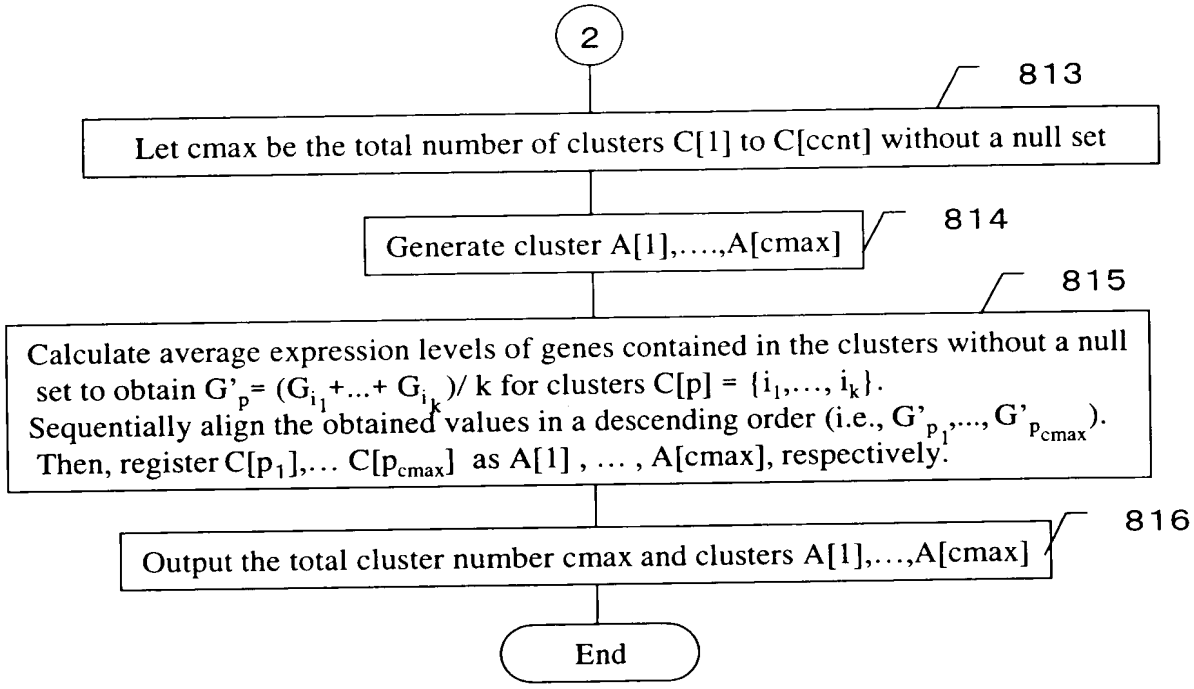

2
813
Let cmax be the total number of clusters C[1] to C[ccnt] without a null set
814
Generate cluster A[1],....,A[cmax]
815
Calculate average expression levels of genes contained in the clusters without a null set to obtain G'p= (Gi1+...+ Gik)/ k for clusters C[p] = {i1,..., ik}.
Sequentially align the obtained values in a descending order (i.e., G'p1,..., G'pcmax).
Then, register C[p1],... C[pcmax] as A[1] , ... , A[cmax], respectively.
816
Output the total cluster number cmax and clusters A[1],...,A[cmax]
End

FIG. 10

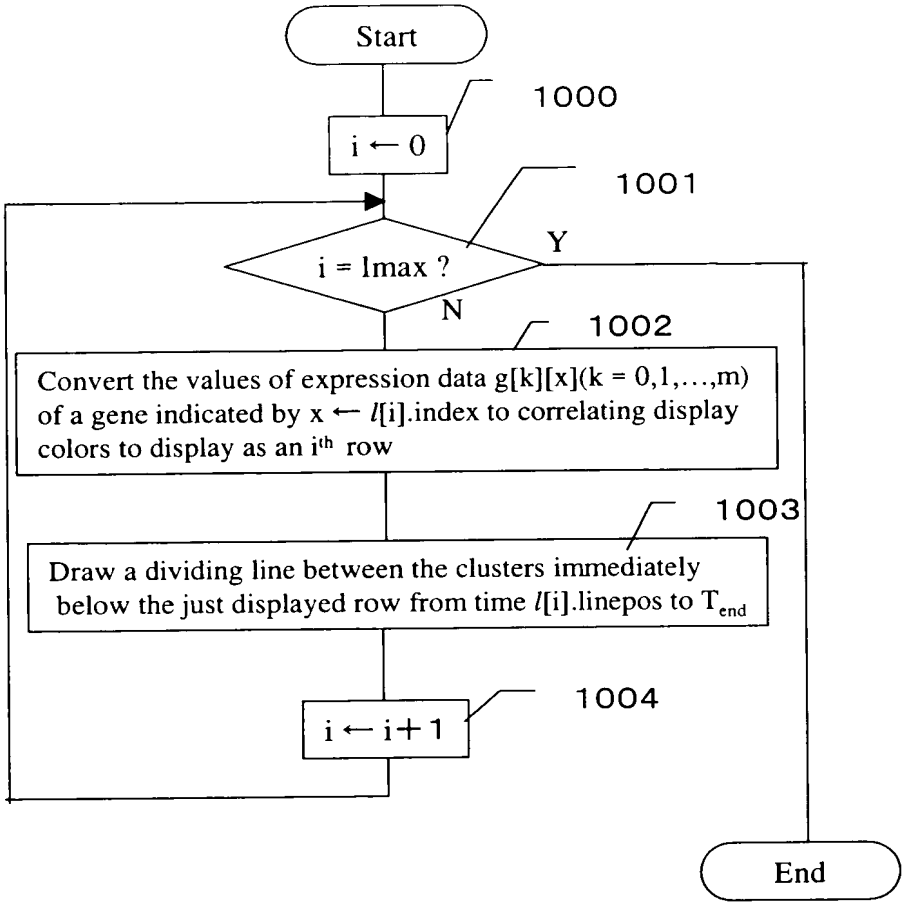

Start
1000
i ← 0
1001
i = lmax ?
Y
N
1002
Convert the values of expression data g[k][x](k = 0,1,...,m) of a gene indicated by x ← l[i].index to correlating display colors to display as an i^th row
1003
Draw a dividing line between the clusters immediately below the just displayed row from time l[i].linepos to T_end
1004
i ← i+1
End

FIG.11

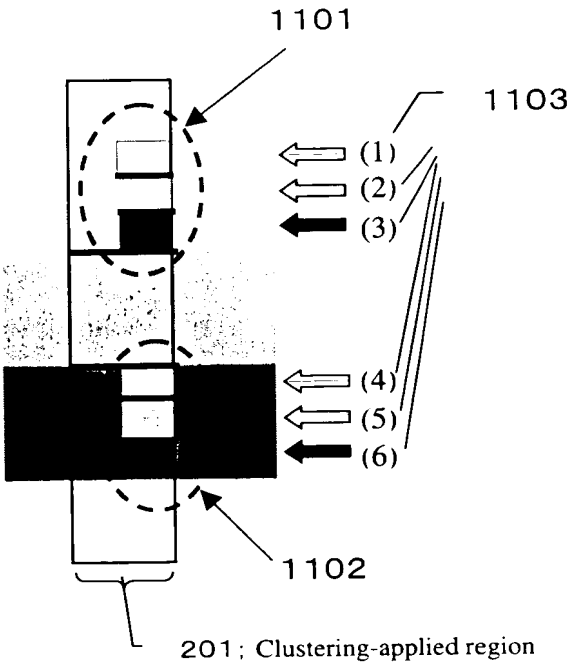


FIG.12

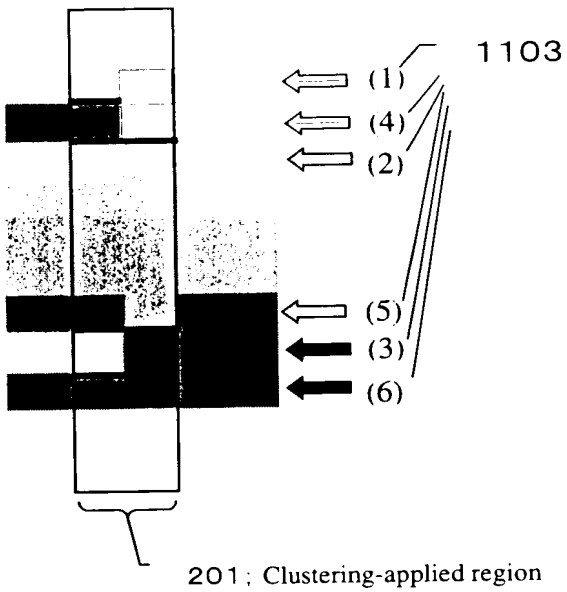


FIG.13

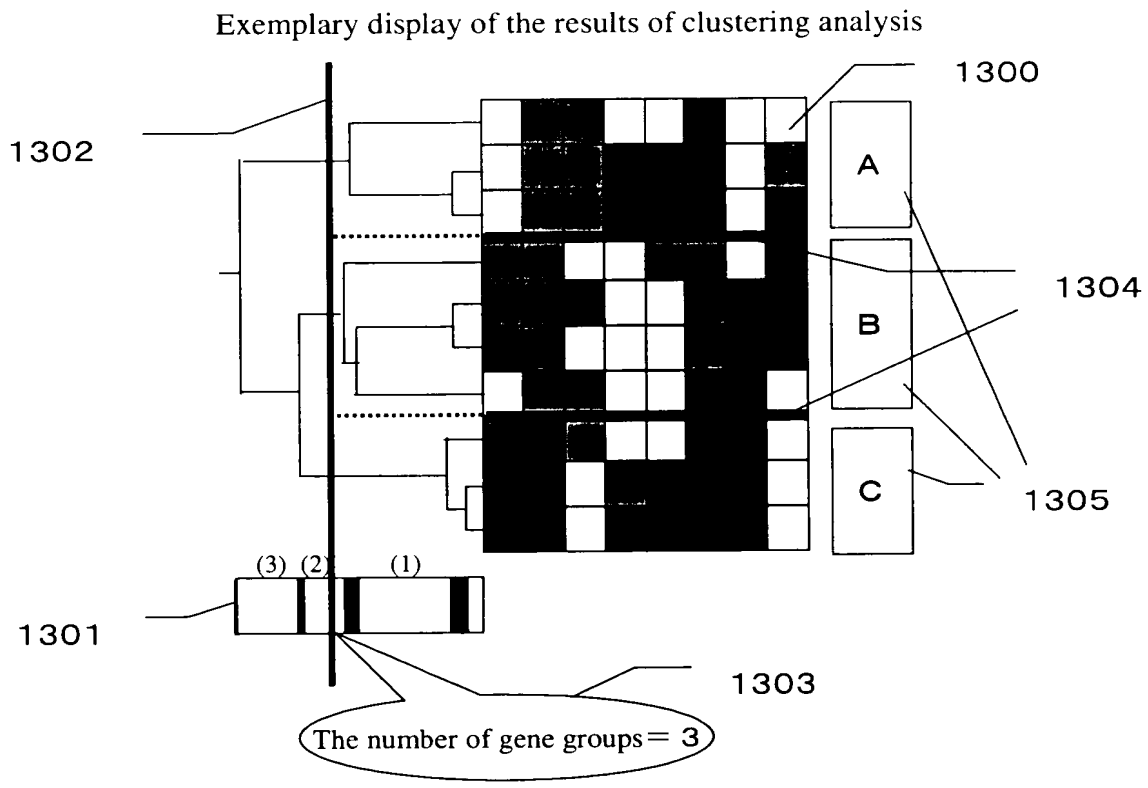

Exemplary display of the results of clustering analysis
1300
1302
A
B
C
1304
1305
(3) (2) (1)
1301
1303
The number of gene groups = 3

FIG.14

Example of gene expression pattern data

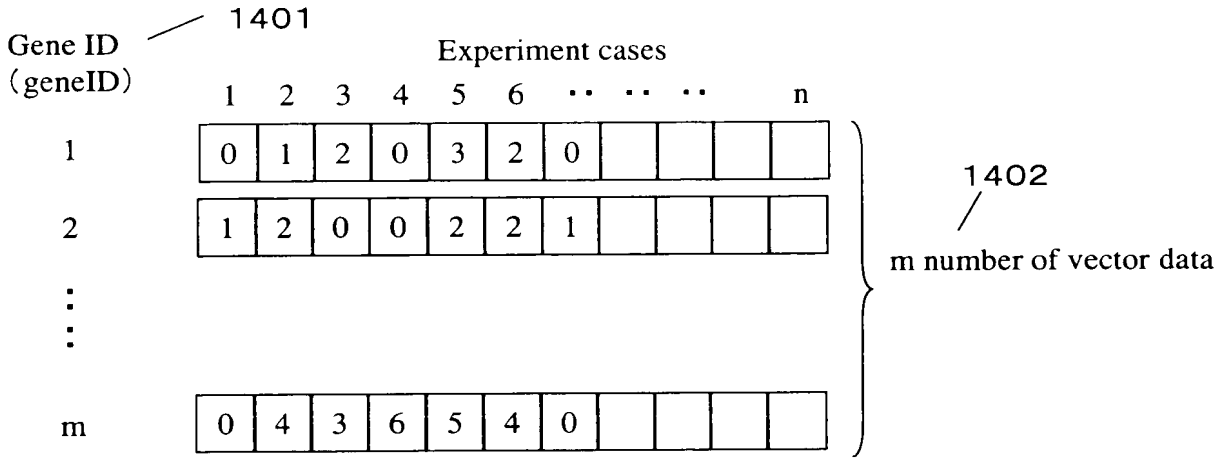


FIG.15

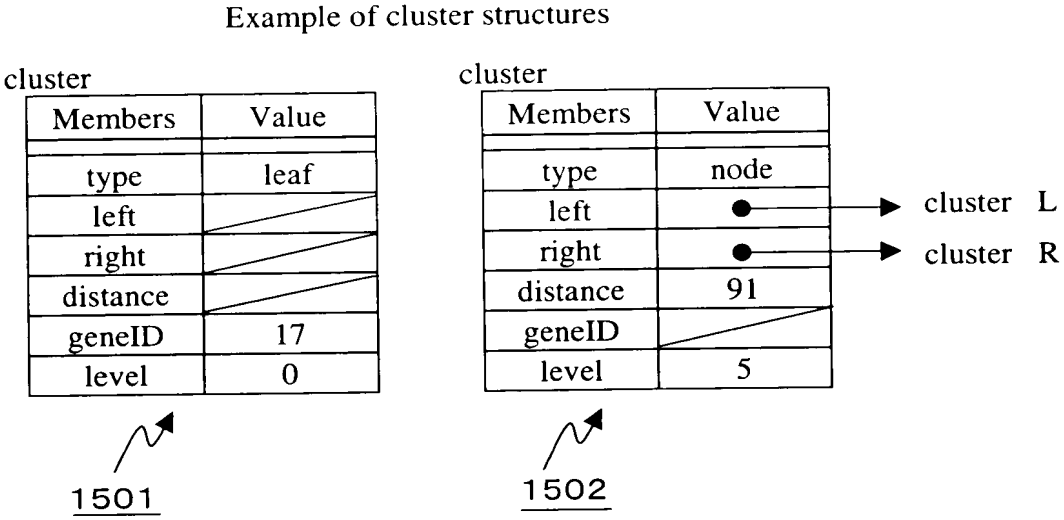

Example of cluster structures
cluster
Members
Value
type
leaf
left
right
distance
geneID
17
level
0
1501
cluster
Members
Value
type
node
left
right
distance
91
geneID
level
5
1502
cluster L
cluster R

FIG.16

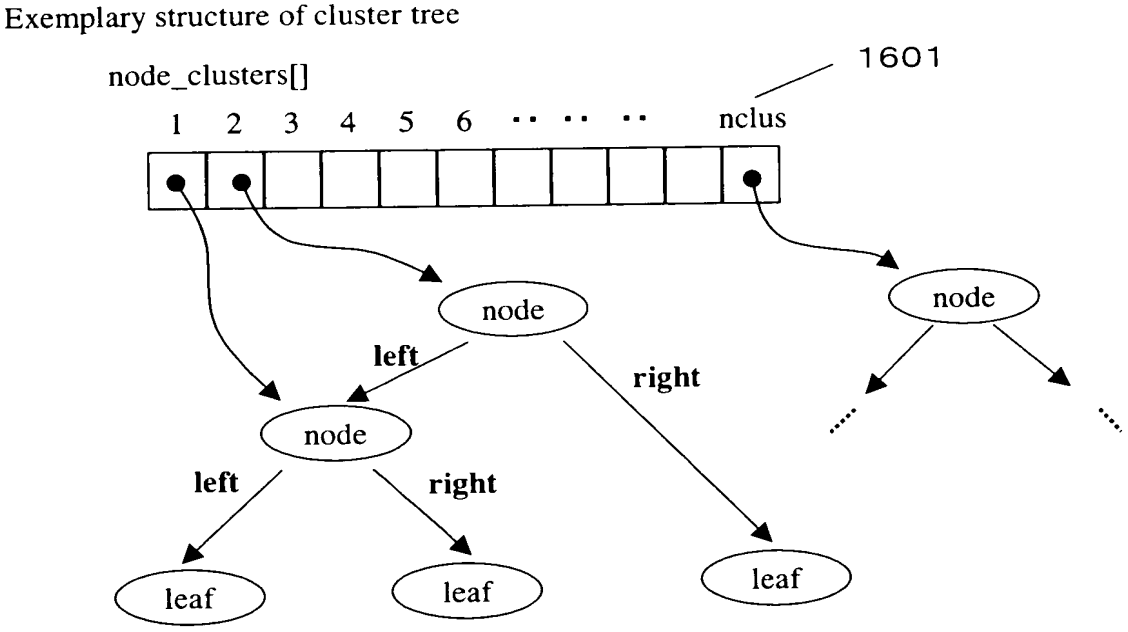

Exemplary structure of cluster tree
node_clusters[]
1601
1 2 3 4 5 6 nclus
node
node
left
right
node
left
right
leaf
leaf
leaf

FIG.17

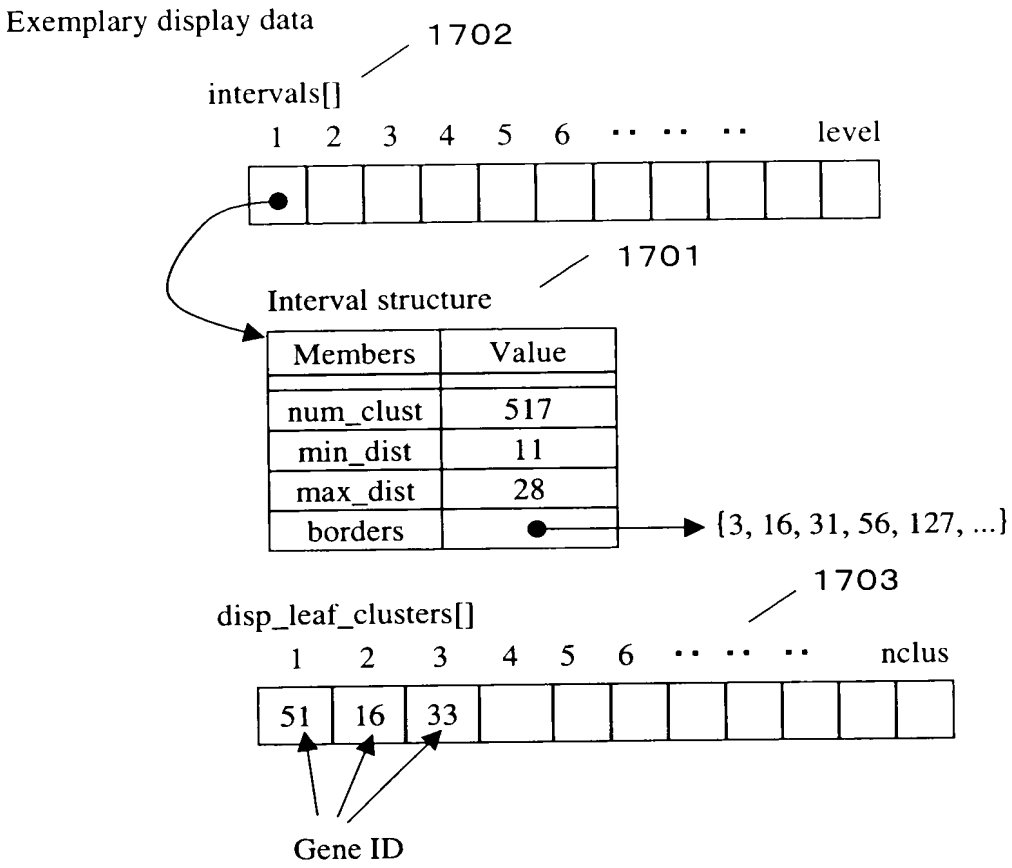

Exemplary display data
1702
intervals[]
1 2 3 4 5 6 level
1701
Interval structure
Members
Value
num_clust
517
min_dist
11
max_dist
28
borders
{3, 16, 31, 56, 127, ...}
1703
disp_leaf_clusters[]
1 2 3 4 5 6 nclus
51
16
33
Gene ID

FIG.18

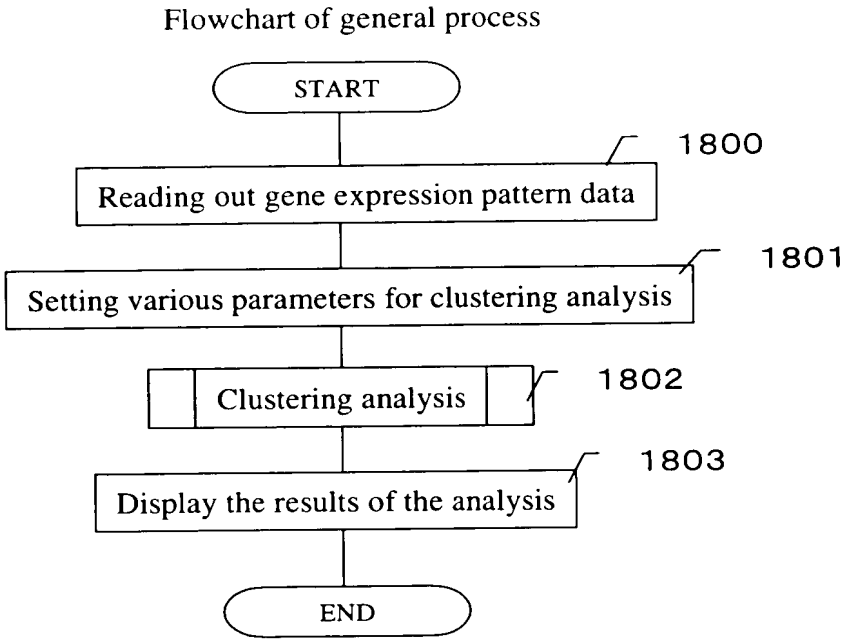

Flowchart of general process
START
1800
Reading out gene expression pattern data
1801
Setting various parameters for clustering analysis
1802
Clustering analysis
1803
Display the results of the analysis
END

FIG.19

Detailed flowchart of clustering analysis (1: Generating cluster tree)

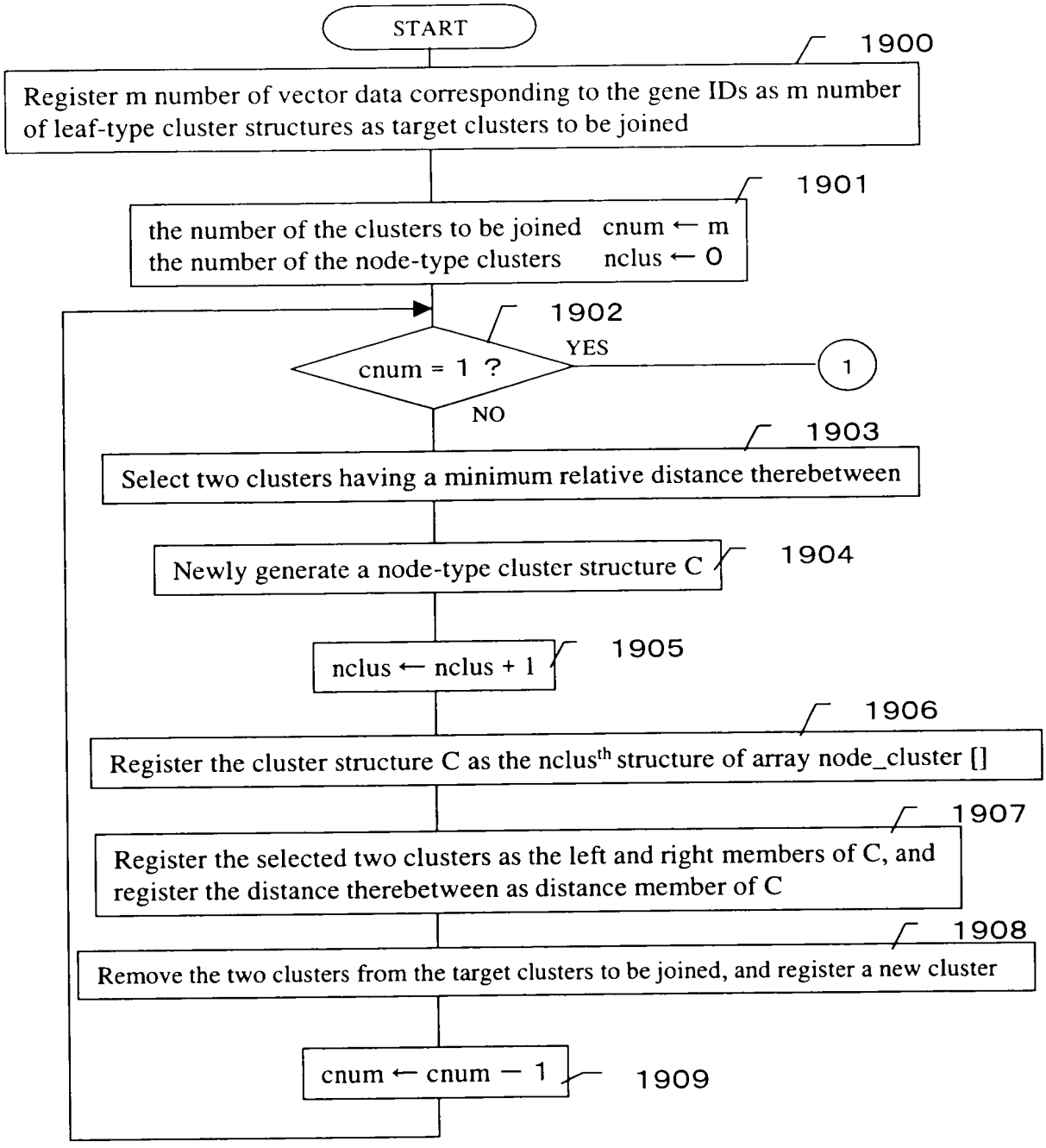


FIG.20

Detailed flowchart of clustering analysis (2: Setting cluster levels)

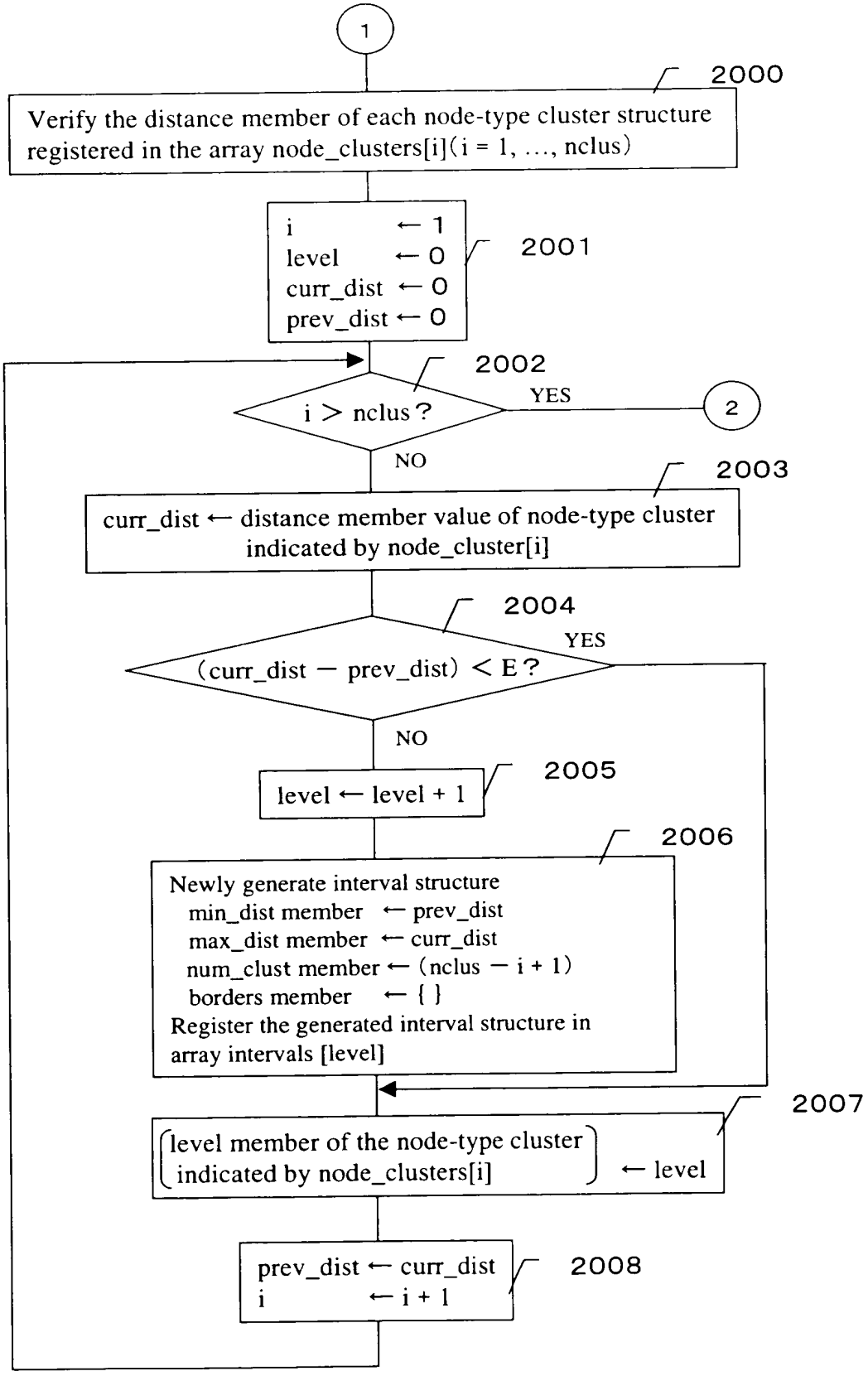


FIG.21

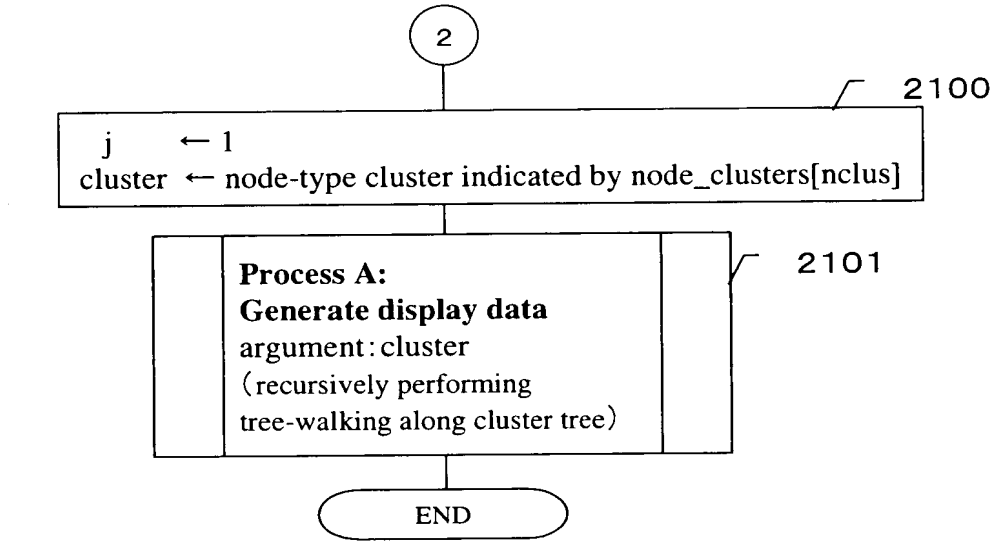

Detailed flowchart of clustering analysis (3: Generating display data)
2
2100
j ← 1
cluster ← node-type cluster indicated by node_clusters[nclus]
Process A:
Generate display data
argument: cluster
(recursively performing
tree-walking along cluster tree)
2101
END

FIG.22

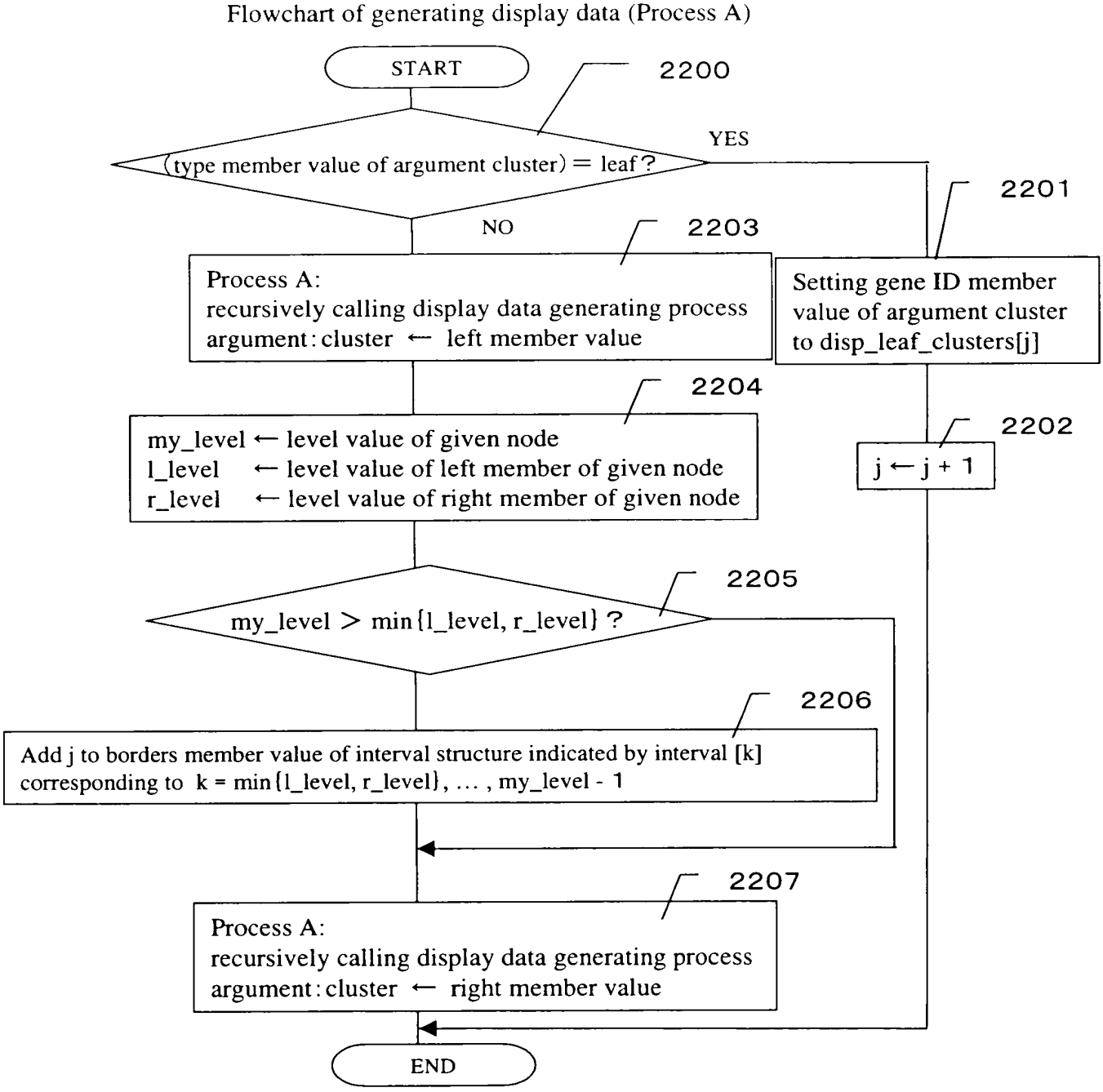

Flowchart of generating display data (Process A)
START
2200
(type member value of argument cluster) = leaf ?
YES
NO
2201
2203
Process A:
recursively calling display data generating process
argument: cluster ← left member value
Setting gene ID member value of argument cluster to disp_leaf_clusters[j]
2204
2202
my_level ← level value of given node
l_level ← level value of left member of given node
r_level ← level value of right member of given node
j ← j + 1
2205
my_level > min{l_level, r_level} ?
2206
Add j to borders member value of interval structure indicated by interval [k]
corresponding to k = min{l_level, r_level}, ... , my_level - 1
2207
Process A:
recursively calling display data generating process
argument: cluster ← right member value
END

FIG.23

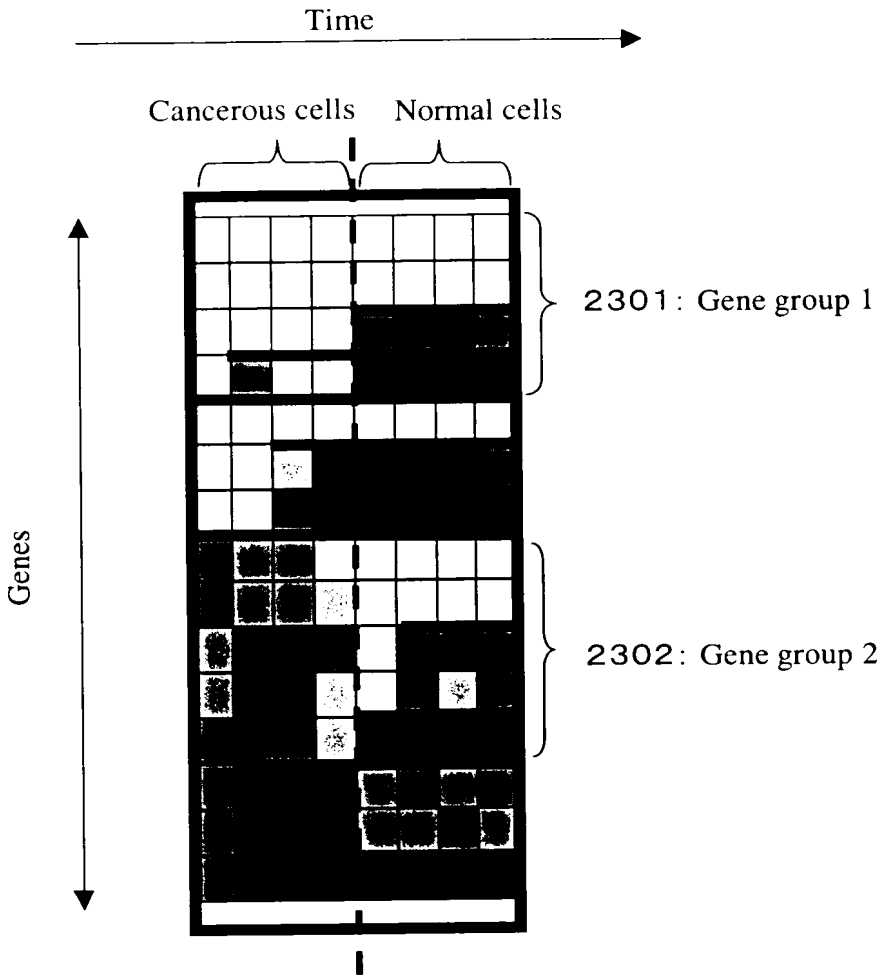

Time
Cancerous cells
Normal cells
Genes
2301 : Gene group 1
2302 : Gene group 2

FIG.24

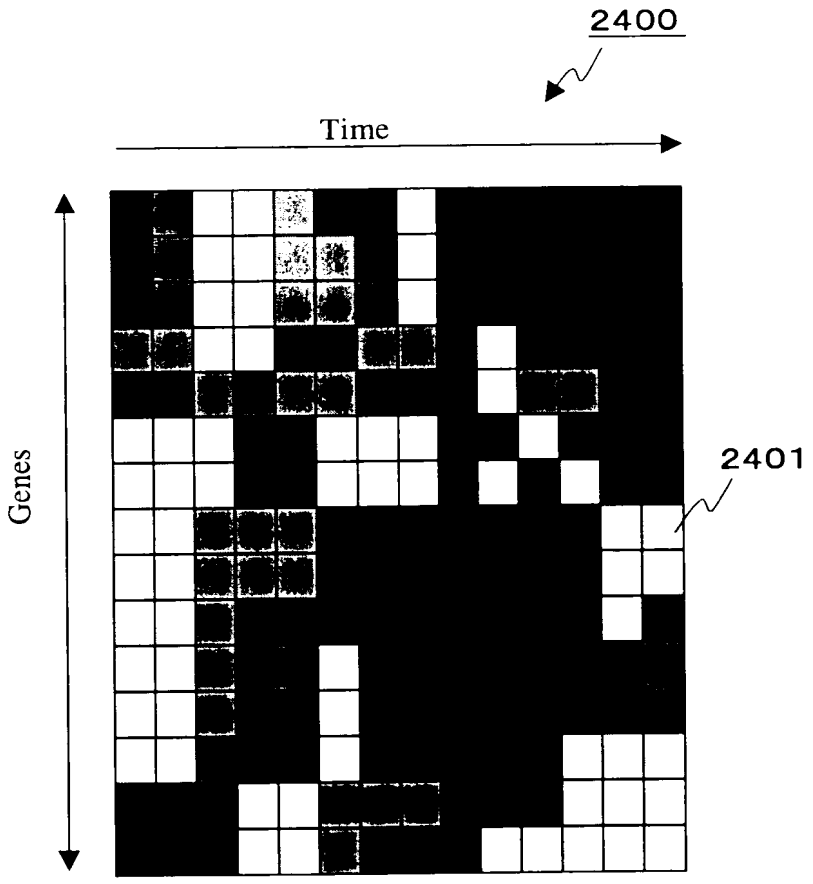


FIG.25

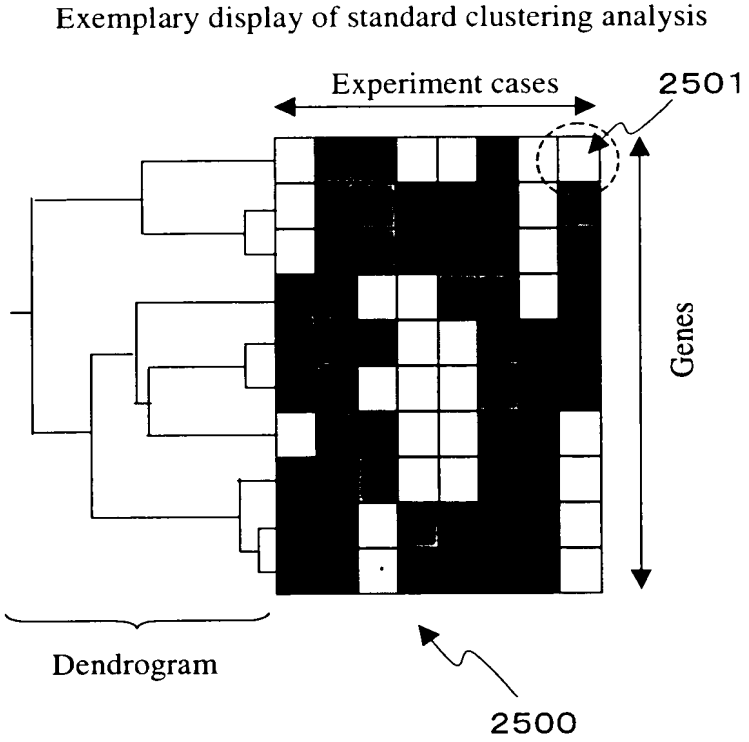
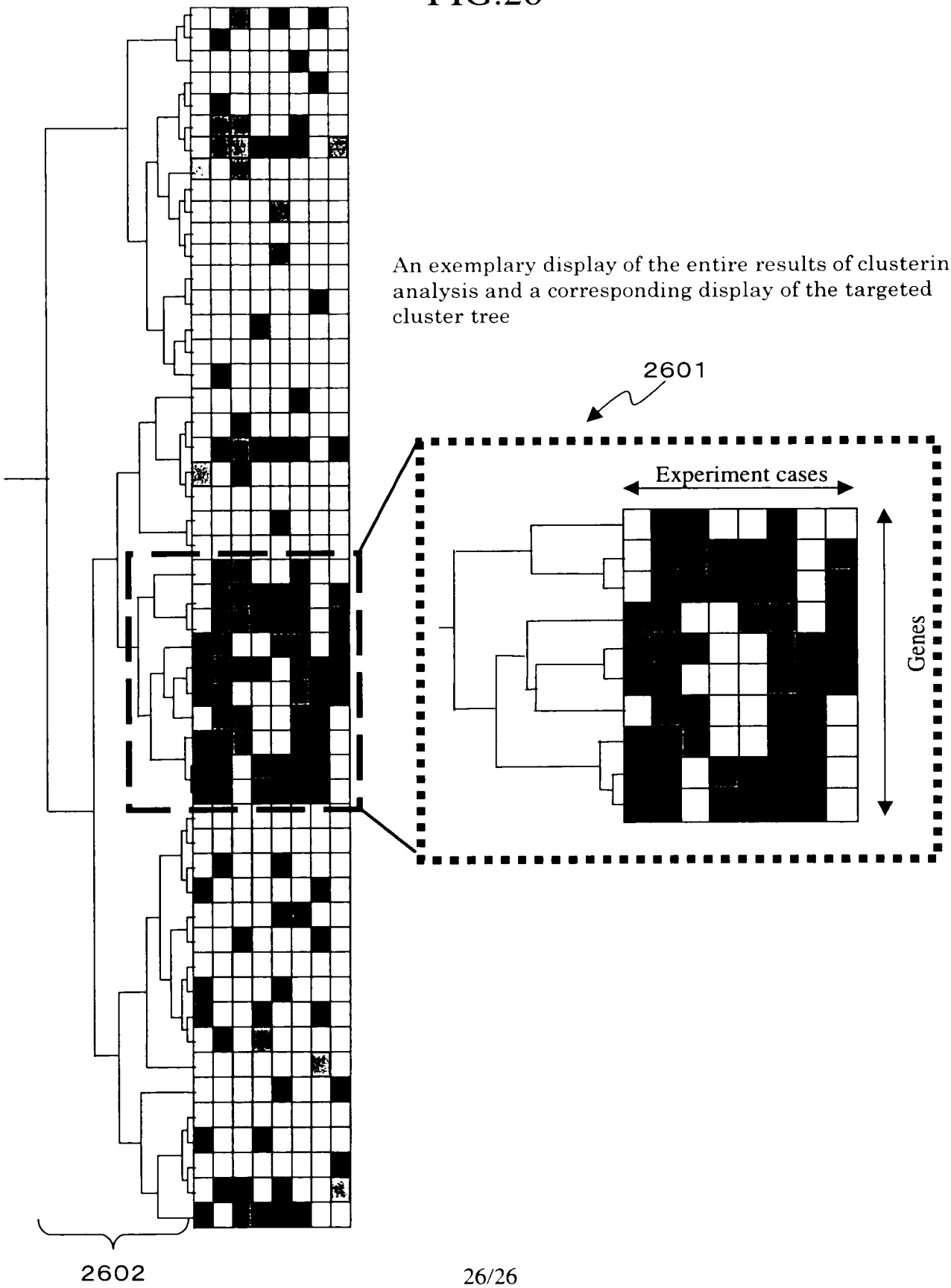

Exemplary display of standard clustering analysis
Experiment cases
2501
Genes
Dendrogram
2500

FIG.26



An exemplary display of the entire results of clustering analysis and a corresponding display of the targeted cluster tree